



Plano de Ensino

Universidade Federal do Espírito Santo

Campus de Goiabeiras

Curso: Arquivologia - Noturno

Departamento Responsável: Departamento de Arquivologia

Data de Aprovação (Art. nº 91):

DOCENTE PRINCIPAL : HENRIQUE MONTEIRO CRISTOVAO

Matrícula: 1727965

Qualificação / link para o Currículo Lattes: <http://lattes.cnpq.br/5035919384923489>

Disciplina: CIÊNCIAS DE DADOS

Código: ARV12967

Período: 2020 / 2

Turma: 1

Pré-requisito:

Carga Horária Semestral: 60

Disciplina: ARV13049 - SISTEMAS GERENCIADORES DE BANCOS DE DADOS (SGDB)

Distribuição da Carga Horária Semestral

Créditos: 2

Teórica

Exercício

Laboratório

30

15

15

Ementa:

Metodologias e técnicas da ciência de dados para armazenamento, análise, tratamento, acessibilidade e visualização de dados.

Objetivos Específicos:

O aluno deverá conhecer as metodologias e técnicas da ciência de dados para armazenamento, análise, tratamento, acessibilidade e visualização de dados.

Conteúdo Programático:

Unidade I: Ciência de Dados

Conceitos básicos da Ciência de Dados

- Descoberta de Conhecimento em Bases de Dados
- Data Warehouse
- Mineração de Dados (Data Mining)
- Big Data
- Metodologias e técnicas da Ciência de Dados

Unidade II: Análise de Redes Complexas na Ciência de Dados

Ciência das Redes

- Fundamentos de Análise de Redes Complexas
- Métricas de Centralidade
- Coesão, Comunidades e Contágio
- Redes Bipartidas

Software para Análise de Redes Complexas

Estudos de casos de análise exploratória em redes complexas

Metodologia:

A disciplina será ministrada subsidiando-se nos parâmetros do EARTE (Ensino-Aprendizagem Remoto Temporário e Emergencial) utilizando-se da plataforma institucional do AVA (Ambiente Virtual de Aprendizagem) e da ferramenta virtual de comunicação/interação Google Meet para os encontros síncronos. O link de acesso às aulas síncronas está disponibilizado na plataforma do AVA. Os alunos deverão acessar o Google Meet utilizando o e-mail institucional.

As aulas serão em todas as terças-feiras de 18h às 22h, sendo que de 19:00h às 20:30h encontros síncronos e o restante do horário aulas assíncronas.

Serão utilizadas sistemáticas adaptadas de sala de aula invertida e de aprendizagem baseada em problemas (PBL) juntamente com aulas síncronas expositivas e dialogadas, com auxílio da ferramenta de comunicação/interação virtual, sempre buscando a melhor compreensão e estimulando a prática de pesquisa bibliográfica por parte dos alunos.

O contexto a ser explorado por meio da aprendizagem baseada em problemas será o desenvolvimento de uma análise exploratória sobre uma base de dados abertos com análise e aplicação de elementos estudados na parte teórica.

Ainda nos encontros síncronos, eventualmente, será necessária a utilização, por parte do aluno, de softwares indicados pelo professor para acompanhamento e execução de ações práticas para o estudo e desenvolvimento de determinados conteúdos.

Para as aulas assíncronas, e como apoio ao ensino-aprendizagem, serão indicados pelo professor: leitura e estudo de bibliografia disponível na internet, visualização de vídeos disponíveis na internet, desenvolvimento de exercícios e resolução de problemas e desafios, execução de atividades práticas em softwares gratuitos disponíveis na internet, e desenvolvimento de trabalhos em grupo apoiados por ferramenta de comunicação/interação virtual.

Ainda nas aulas assíncronas, e conforme necessidade e disponibilidade da turma ou de um grupo de alunos, será possível a realização de pequenos encontros para a retirada de dúvidas, apoio ao desenvolvimento de atividades práticas, orientações e discussões para encaminhar possíveis soluções na resolução de problemas por intermédio da interação entre professor-aluno-aluno apoiada pela ferramenta virtual.

OBSERVAÇÕES

1. No momento da aula síncrona no Google Meet, será feita chamada para registro de presença.
2. As atividades síncronas e assíncronas poderão ser gravadas para utilização restrita aos fins a que se destina esta disciplina específica, facultando-se ao aluno seu direito de não ter sua imagem gravada ou filmada, mediante expressa manifestação.
3. Não autorizamos a divulgação em qualquer mídia de nossa imagem, voz ou produto do trabalho apresentado em aulas síncronas ou assíncronas em qualquer material que não seja especificamente para fins educacionais, culturais e formativos da disciplina.
4. É VEDADA a reprodução das aulas gravadas sem a autorização expressa do professor e dos alunos participantes.
5. Todos os materiais disponibilizados durante o curso são exclusivamente para o uso didático na disciplina e para fins educacionais, sendo vedada a sua utilização para qualquer outra finalidade, sob as penas legais.
6. Alterações poderão ser realizadas a partir das necessidades ocorridas ao longo do desenvolvimento das aulas e atividades, em consonância com as orientações do docente.

Critérios / Processo de avaliação da Aprendizagem :

A avaliação da aprendizagem dos alunos irá considerar os aspectos qualitativos e quantitativos de obtenção da nota final e por meio da promoção de uma pedagogia da inclusão e obedecendo os objetivos e os conteúdos explicitados no presente plano de ensino.

Será respeitado o prazo mínimo de uma semana de antecedência para marcação de atividade avaliativa, sendo que as atividades avaliativas assíncronas terão um prazo de entrega mínimo de 48 horas para a sua realização.

As avaliações serão divididas em cinco ações, a serem desenvolvidas em momentos assíncronos: dois questionários sobre a parte teórica disponíveis no AVA, com valor total de 2 pontos; dois trabalhos práticos desenvolvidos com apoio de softwares indicados pelo professor, de forma individual ou em grupo, com valor total de 2 pontos; Um trabalho prático final desenvolvido de forma individual ou em grupo, com valor total de 6 pontos.

O texto referente ao trabalho final será desenvolvido e entregue em momento assíncrono.

Haverá apresentação dos trabalhos finais em encontros síncronos nas duas últimas semanas de aula.

A média parcial será calculada pela soma das cinco avaliações descritas.

Se a média parcial for maior ou igual a 7,0 o aluno será aprovado, caso contrário, o aluno terá oportunidade de fazer uma prova final com pontuação de 0 a 10. Nesse caso, a média final será:
(média parcial + nota da prova final) / 2, e a aprovação será obtida com um valor maior ou igual 5.

Bibliografia básica:

CARVALHO, L. A. V. – Data Mining: A mineração de dados no marketing, medicina, engenharia e administração. São Paulo: Érica, 2001. FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMITH, P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In Proceedings of the Second International Conference on Data Mining and Knowledge Discovery, AAAI Press, Menlo Park, US; 1996. Disponível em [kdd96-014.pdf](#). Acesso em Maio de 2016. REZENDE, Solange Oliveira. Mineração de Dados. Anais do V Encontro Nacional de Inteligência Artificial, Sociedade Brasileira de Computação. ISBN: 8576690330. São Leopoldo RS, 25 a 29 de julho de 2005. Disponível em [0102.pdf](#). Acesso em Maio de 2016. SILVA, Leandro. A.; SILVA Luciano. Fundamentos de Mineração de Dados Educacionais. Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 3º Congresso Brasileiro de Informática na Educação (CBIE 2014) Workshops (WCBIE 2014). Disponível em [cbie.wcbie.2014.568](#). Acesso em Maio de 2016.

Bibliografia complementar:

BAKER, R.; ISOTANI, S.; DE CARVALHO, A. Mineração de Dados Educacionais: Oportunidades para o Brasil. Revista Brasileira de Informática na Educação, Volume 19, Número 2, 2011. Disponível em [67](#). Acesso em Maio de 2016. CHEN, Hsinchun; CHIANG, Roger HL; STOREY, Veda C. Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly, v. 36, n. 4, p. 1165-1188, 2012. Disponível em . Acessado em Junho de 2015. CURTY, Renata Gonçalves; SERAFIM, Jucenir da Silva. A formação em ciência de dados: uma análise preliminar do panorama estadunidense. Revista Informação e Informação. DOI:10.5433/1981-8920.2016v21n2p307. Londrina, v. 21, n. 2, p. 307–328. Disponível em: <http://www.uel.br/revistas/informacao/>. Acesso em jun de 2019. FERRARI, Luciana Itida. Uma Metodologia para Extração de Informação sobre o Sistema Imunológico. Tese (Doutorado). Coordenação do Programa de pósgraduação em Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro (COPPE/UFRJ), RJ, 2008. LOTT, Yuri Monnerat and CIANCONI, Regina de Barros. Vigilância e privacidade, no contexto do big data e dados pessoais: análise da produção da Ciência da Informação no Brasil. *Perspect. ciênc. inf.* [online]. 2018, vol.23, n.4, pp.117-132. ISSN 1413-9936. Disponível em: <http://dx.doi.org/10.1590/1981-5344/3313>. Acesso em jun de 2019. PIATETSKY-SHAPIRO, G. FAYYAD, U. An Introduction to SIGKDD and A Reflection on the Term 'Data Mining'. SIGKDD Explorations Volume 13, Issue 2, Page 102, 2011. Disponível em . Acesso em Maio de 2016.

Cronograma:

Observação:

OUTRAS BIBLIOGRAFIAS

ALBERT-LÁSZLÓ BARABÁSI. Linked: a nova ciência dos networks. Tradução Jonas Pereira dos Santos. 1. ed. São Paulo, SP: Leopardo, 2009.

WITTEN, I.H; FRANK, E. Data Mining Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 3º ed., 2011.

BIBLIOGRAFIA ADAPTADA AO EART (Art. 5º, § 2º, da Resolução CEPE nº 30/2020)

CRISTOVÃO, H. Análise de redes complexas na Ciência de Dados. Página web, 2020. Disponível em <http://bit.ly/henriquecristovao-arccd>.

CRISTOVÃO, H. Estudos de casos de análise exploratória sobre bases de dados abertos. Página web, 2020. Disponível em <http://bit.ly/henrique-cristovao-ecae>.

CURTY, Renata Gonçalves; SERAFIM, Jucenir da Silva. A formação em ciência de dados: uma análise preliminar do panorama estadunidense. Revista Informação e Informação. DOI:10.5433/1981-8920.2016v21n2p307. Londrina, v. 21, n. 2, p. 307-328. Disponível em: <http://www.uel.br/revistas/informacao/>. Acesso em jun de 2019.

FRANCO, Augusto de. A rede. São Paulo: Escola de Redes, 2012. Disponível em: <http://pt.slideshare.net/augustodefranco/fluzz-srie-completa>.

LOTT, Yuri Monnerat; CIANCONI, Regina de Barros. Vigilância e privacidade, no contexto do big data e dados pessoais: análise da produção da Ciência da Informação no Brasil. *Perspect. ciênc. inf.* [online]. 2018, vol.23, n.4, pp.117-132. ISSN 1413-9936. Disponível em: <http://dx.doi.org/10.1590/1981-5344/3313>. Acesso em jun de 2019.

TOMAÉL, Maria Inês. Redes de conhecimento. DataGramaZero - Revista de Ciência da Informação, v. 9, n. 2, abr. 2008. Disponível em: http://www.datagramazero.org.br/abr08/Art_04.htm. Acesso em: 31 mar. 2016.

CRONOGRAMA

Início do período letivo: 01/02/2021

Término do período letivo: 15/05/2021

Provas finais: 17 a 21/05/2020

Aulas síncronas às quintas-feiras das 19h às 21h.

Os demais horários, de 18h às 19h e 21h às 22h, serão destinados para as atividades assíncronas.

As datas das aulas seguem abaixo, juntamente com o conteúdo planejado, podendo ocorrer eventualmente alterações mediante consenso com a turma.

11-fevereiro: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Apresentação da metodologia da disciplina na modalidade EARTE e a proposta de sala de aula invertida e a aprendizagem baseada em problemas (PBL).

Proposta para a PBL: desenvolvimento de uma análise exploratória sobre uma base de dados abertos com análise e aplicação de elementos estudados na parte teórica.

Introdução à Ciência de Dados.

18-fevereiro: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Elementos fundamentais de Ciência de Dados.

25-fevereiro: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Conhecer algumas metodologias e técnicas da Ciência de Dados

Interfaces entre Arquivologia, Ciência de Dados, Mineração de Dados, Datawarehousing

04-março: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Princípios de Aprendizado de Máquina e de Inteligência Artificial, e suas aplicações.

O processo do KDD e suas etapas.

11-março: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Fundamentos de Análise de Redes Complexas.

Métricas de Centralidade.

Coesão, Comunidades e Contágio.

18-março: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Software para Análise de Redes Complexas.

Estudo de caso 1.

25-março: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Estudo de caso 2.

01-abril: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Estudo de caso 3.

08-abril: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Redes Bipartidas.

Continuação do Estudo de caso 3

15-abril: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Apresentação da proposta de trabalho final segundo a PBL.

Estudo de caso 4.

22-abril: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Continuação do estudo de caso 4.

Orientações e retirada de dúvidas direcionadas ao desenvolvimento do trabalho final.

29-abril: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Estudo de caso 5.

Apresentação pelos grupos do contexto e base de dados escolhida e as pretensões para a análise exploratória.

06-maio: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.

Apresentação do trabalho final pelos grupos.

13-maio: Aula síncrona de 19:00 às 21:00 e aula assíncrona de 18:00 às 19:00 e de 21:00 às 22:00.
Apresentação do trabalho final pelos grupos.

20-maio:
Prova Final.